

AMENDMENTS TO THE SPECIFICATION:

With regards to the Examiners objection to the specification:

Please replace the paragraph on page 4, lines 8-11 with:

The problem of inconsistent naming is exemplified by considering the chemical names that have been applied to the drug ~~Valium~~TM VALIUM^R (Valium is a registered trademark of Roche Products Inc.), the chemical structure of which is shown in Fig. 1. A list of some of the correct and incorrect names for ~~Valium~~TM VALIUM^R that are found in the chemical and patent literature are shown in Table 1.

Please replace the paragraph on page 4, line 22 to page 5, line 10 with:

Additionally, in the case of pharmaceuticals, the names of compounds of interest often change over time as compounds become commercialized. This has led to the frequent use of trade names or generic names in the scientific literature or in medical databases, which are not reflected retrospectively in the various IP databases. This has made it difficult to perform text searching for certain pharmaceuticals in the patent literature using commonly accepted phrases or definitions. For example, one cannot simply type in the search term "aspirin" or "~~Valium~~TM" "VALIUM^R" into any of the IP databases and find the pertinent patents for those chemical substances. The problem is further exacerbated by the fact that different brand names are often used in different countries to address language considerations of the different geographical areas. In fact, there are as many as 149 different names that have been employed in the literature for the drug ~~Valium~~TM VALIUM^R, a number of which are illustrated in Table 2.

Please replace the paragraph on page 5, line 12 with:

Table 2 - Some of the trade names used to refer to ~~Valium~~TM VALIUM^R

Please replace paragraph on page 8, line 16 with:

Fig. 1 shows the chemical structure of ~~Valium~~TM VALIUM^R;

Please replace the paragraph on page 8, line 21 with:

Fig. 3 shows various chemical substructures parsed from the chemical name for ~~Valium~~TM VALIUM^R;

Please replace the paragraph on page 11, lines 4-6 with:

Thus, while the numerous variations in the name of ~~Valium~~TM VALIUM^R in Tables 1 and 2 are too extensive for a text search to be helpful, a search for the fragments by structure is much more likely to be successful.

Please replace the paragraph on page 11, lines 8-16 with:

In mining information from text documents, such as patents and technical articles, it is critical that long multi-word organic chemical nomenclatures be recognized properly so they can be grouped as single logical entities and correctly indexed. In the above-referenced commonly assigned U.S. Patent Application S.N. 10/670,675 the inventors Coden and Cooper previously described a system and method for grouping such nomenclature into logical entities without the need to provide large chemical dictionaries. This invention makes use of a search engine, such as one known as a ~~JuruXML~~TM JURUXML^R search engine available from the assignee of this patent application, and a table of substructure names and connectivity. Such a table could, for example, be stored in a relational database such as one known as DB2TM, also available from the assignee of this patent application.

Please replace the paragraph on page 17, lines 6-11 with:

Fig. 6A describes indexing a collection of documents. Each document is read in from a file (block 600) and indexed (block 601) in a conventional manner using a search engine, such as the ~~JuruXML~~TM JURUXML^R search engine. In the presently preferred embodiment the algorithm described in the commonly assigned U.S. Patent Application S.N. 10/670,675 is then used to identify organic chemical names (block 602). Each organic chemical name is separated into sub-tokens, separated by, for example, hyphens, spaces and parenthesis (block 603).